

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

-Tue Apr 17 15:46:22 2001

AD-6/23/1997
✓ 6/22/1998

us-09-446-543a-73.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:39:49 ; Search time 61.54 Seconds
(without alignments)
6.868 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPMWXXRGIRPVGRFX 22

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	96.3	20	3	US-09-105-678A-46
2	105	96.3	21	3	US-09-105-678A-47
3	105	96.3	22	3	US-09-105-678A-48
4	105	96.3	31	3	US-09-105-678A-9
5	105	96.3	31	3	US-09-105-678A-43
6	105	96.3	32	3	US-09-105-678A-44
7	105	96.3	33	3	US-09-105-678A-45
8	104	95.4	20	3	US-09-105-678A-34
9	104	95.4	21	3	US-09-105-678A-35
10	104	95.4	21	3	US-09-105-678A-41
11	104	95.4	22	3	US-09-105-678A-36
12	104	95.4	22	3	US-09-105-678A-42
13	104	95.4	31	3	US-09-105-678A-7
14	104	95.4	31	3	US-09-105-678A-8
15	104	95.4	31	3	US-09-105-678A-31
16	104	95.4	31	3	US-09-105-678A-37
17	104	95.4	31	4	US-09-172-353-4
18	104	95.4	32	3	US-09-105-678A-38
19	104	95.4	32	3	US-09-105-678A-39
20	104	95.4	33	3	US-09-105-678A-33
21	104	95.4	33	3	US-09-105-678A-39
22	104	95.4	33	3	US-09-105-678A-28
23	103	94.5	21	3	US-09-105-678A-30
24	98	89.9	19	3	US-09-105-678A-29
25	93	85.3	29	3	US-08-118-270-71
26	44	40.4	349	1	PCR-US93-08528-71
27	44	40.4	349	5	Sequence 71, Appl

28	41	37.6	547	1	US-08-083-948-8	Sequence 8, Appl1
29	41	37.6	547	1	US-08-393-785-8	Sequence 8, Appl1
30	41	37.6	547	1	US-08-475-694-8	Sequence 8, Appl1
31	41	37.6	547	1	US-08-712-057-8	Sequence 8, Appl1
32	39	35.8	162	4	US-08-992-176-9	Sequence 9, Appl1
33	39	35.8	555	2	US-08-982-232-7	Sequence 7, Appl1
34	39	35.8	555	2	US-08-982-232-14	Sequence 14, Appl1
35	38.5	35.3	555	2	US-08-453-702A-98	Sequence 98, Appl1
36	38.5	35.3	831	1	US-08-073-384C-5	Sequence 5, Appl1
37	38.5	35.3	831	1	US-08-254-359A-5	Sequence 5, Appl1
38	38.5	35.3	831	1	US-08-483-043-5	Sequence 5, Appl1
39	38.5	35.3	831	1	US-08-481-238-5	Sequence 5, Appl1
40	38.5	35.3	831	2	US-08-471-066B-5	Sequence 5, Appl1
41	38.5	35.3	831	2	US-08-484-956-5	Sequence 5, Appl1
42	38.5	35.3	831	2	US-08-757-653-5	Sequence 5, Appl1
43	38.5	35.3	831	2	US-08-599-491-5	Sequence 5, Appl1
44	38.5	35.3	831	2	US-08-756-386-5	Sequence 5, Appl1
45	38.5	35.3	831	2	US-08-823-516-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1

US-09-105-678A-46
Sequence 46, Application US/09105678A
Patent No. 6103862

GENERAL INFORMATION:
APPLICANT: Suenaga, Masato

APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko

APPLICANT: Mishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street

CITY: Boston
STATE: MA

COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998

PRIOR APPLICATION DATA:
APPLICATION NUMBER: (JP 172118/1997)

FILING DATE: (27-JUN-1997)
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026

TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 46:

SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids

TYPE: amino acid
STRANDEDNESS:

TOPOLOGY: linear
MOLECULE TYPE: peptide

US-09-105-678A-46

Query Match 96.3%; Score 105; DB 3; Length 20;
Best Local Similarity 90.0%; Pred. No. 1.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-9

Query Match 96.3%; Score 105; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 2.4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 5
US-09-105-678A-43
Sequence 43, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-43

QY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 6
US-09-105-678A-44
Sequence 44, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27, 026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 44:
SEQUENCE CHARACTERISTICS:
LENGTH: 32 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-44

Query Match 96.3%; Score 105; DB 3; Length 32;
Best Local Similarity 90.0%; Pred. No. 2.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXRGIRPVGRF 20
|||||
Db 12 TPDINPAMYASRGIRPVGRF 31

RESULT 7
US-09-105-678A-45
Sequence 45, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-45

Query Match 96.3%; Score 105; DB 3; Length 33;
Best Local Similarity 90.0%; Pred. No. 2.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
||||||| |||||||
DB 12 TPDINPAMYASRGIRPVGRF 31

RESULT 8
US-09-105-678A-34
Sequence 34, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:

NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-34

Query Match 95.4%; Score 104; DB 3; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.2e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TPDINPAMYXXRGIRPVGRF 20
||||||| |||||||
DB 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 9
US-09-105-678A-40
Sequence 40, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-40

Query Match 95.4%; Score 104; DB 3; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.2e-11;

	Matches	18, Conservative	0;	Mismatches	2;	Indels	0;	Gaps	0;
QY	1	TPDINPAWYXXRGIRPVGRF	20						
Db	1	TPDINPAWYTGRCIRPVGRF	20						

RESULT 10
US-09-105-

```

: Sequence 35 Application US/09105678A
: Patent No. 6103882
:
: GENERAL INFORMATION:
: APPLICANT: Suenaga, Masato
: APPLICANT: Moriya, Takeo
: APPLICANT: Tanaka, Yoko
: APPLICANT: Nishimura, Osamu
: TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
: NUMBER OF SEQUENCES: 52
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
: STREET: 130 Water Street
: CITY: Boston
: STATE: MA
:
: COUNTRY: USA
: ZIP: 02109
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/105,678A
: FILING DATE: 26-JUN-1998
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 172118/1997
: FILING DATE: 27-JUN-1997
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Conlin, David G.
: REGISTRATION NUMBER: 27,026
: REFERENCE/DOCKET NUMBER: 48466-342
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-523-3400
: TELEFAX: 617-523-6440
:
: INFORMATION FOR SEQ ID NO: 35:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
:
: MOLECULE TYPE: peptide
:
: US-09-105-678A-35
:
:
: Query Match 95.4%; Score 104; DB 3; Length 21;
: Best Local Similarity 90.0%; Pred. NO. 2.3e-11;
: Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0.
:
: QY 1 TPDIHPAWYXXKGRIPVGRF 20
: ||||||| |||||||
: Db 1 TPDIHPAWYAGRGIRPVGRF 20
:
:
: RESULT 11
: US-09-105-678A-41
: Sequence 41, Application US/09105678A
: Patent No. 6103882
:
: GENERAL INFORMATION:
: APPLICANT: Suenaga, Masato
: APPLICANT: Moriya, Takeo
: APPLICANT: Tanaka, Yoko
: APPLICANT: Nishimura, Osamu
: TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
: NUMBER OF SEQUENCES: 52
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
: STREET: 130 Water Street
: CITY: Boston
: STATE: MA
:
: COUNTRY: USA
: ZIP: 02109
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/105,678A
: FILING DATE: 26-JUN-1998
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 172118/1997
: FILING DATE: 27-JUN-1997
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Conlin, David G.
: REGISTRATION NUMBER: 27,026
: REFERENCE/DOCKET NUMBER: 48466-342
:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-523-3400
: TELEFAX: 617-523-6440
:
: INFORMATION FOR SEQ ID NO: 35:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 21 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
:
: MOLECULE TYPE: peptide
:
: US-09-105-678A-35
:

```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
3 STREET: 130 Water Street
4 CITY: Boston
5 STATE: MA
6 COUNTRY: USA
7 ZIP: 02109

COMPUTER READABLE FORM:

```

1      MEDIUM TYPE: Floppy disk
2      COMPUTER: IBM PC compatible
3      OPERATING SYSTEM: PC-DOS/MS-DOS
4      SOFTWARE: Patentin Release #1.0, Version #1.30
5      CURRENT APPLICATION DATA:
6      APPLICATION NUMBER: US/09/105,678A
7      FILING DATE: 26-JUN-1998
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: JP 172118/1997
10     FILING DATE: 27-JUN-1997
11     ATTORNEY/AGENT INFORMATION:
12     NAME: Conlin, David G.
13     REGISTRATION NUMBER: 27,026
14     REFERENCE/DOCKET NUMBER: 48466-342
15     TELECOMMUNICATION INFORMATION:
16     TELEPHONE: 617-523-3400
17     TELEFAX: 617-523-6440
18     INFORMATION FOR SEQ ID NO: 41:
19     SEQUENCE CHARACTERISTICS:
20     LENGTH: 21 amino acids
21     TYPE: amino acid
22     STRANDEDNESS:
23     TOPOLOGY: linear
24     MOLECULE TYPE: peptide
25     US-09-105-678A-41

```

Query Match	95.48;	Score 104;	DB 3;	Length 21;
Best Local Similarity	90.08;	Pred. No. 2.3e-11;		
Matches 18; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY 1 TPDINPAWYXXRGIRPVGRF 20
||||| |
Db 1 TPDINPAWYTGGRIRPVGRF 20

RESULT 12

US-09-105-678A-36
Sequence 36, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997

ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-36

Query Match 95.4%; Score 104; DB 3; Length 22;
Best Local Similarity 90.0%; Pred. No. 2.4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXGIRPVGRF 20
DB 1 TPDINPAMYAGRGIRPVGRF 20

RESULT 13

US-09-105-678A-42
Sequence 42, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 22 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-42

Query Match 95.4%; Score 104; DB 3; Length 22;

Best Local Similarity 90.0%; Pred. No. 2.4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXGIRPVGRF 20
DB 1 TPDINPAMYTGIRPVGRF 20

RESULT 14

US-09-105-678A-7
Sequence 7, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
STREET: 130 Water Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,678A
FILING DATE: 26-JUN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 172118/1997
FILING DATE: 27-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Conlin, David G.
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 48466-342
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 31 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-105-678A-7

Query Match 95.4%; Score 104; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 3.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TPDINPAMYXXGIRPVGRF 20
DB 12 TPDINPAMYAGRGIRPVGRF 31

RESULT 15

US-09-105-678A-8
Sequence 8, Application US/09105678A
Patent No. 6103882
GENERAL INFORMATION:
APPLICANT: Suenaga, Masato
APPLICANT: Moriya, Takeo
APPLICANT: Tanaka, Yoko
APPLICANT: Nishimura, Osamu
TITLE OF INVENTION: METHOD OF PRODUCING A 19P2 LIGAND


```

:      NUMBER OF SEQUENCES: 52
:      CORRESPONDENCE ADDRESS:
:      ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
:      STREET: 130 Water Street
:      CITY: Boston
:      STATE: MA
:      COUNTRY: USA
:      ZIP: 02109
:      COMPUTER READABLE FORM:
:      MEDIUM TYPE: Floppy disk
:      COMPUTER: IBM PC compatible
:      OPERATING SYSTEM: PC-DOS/MS-DOS
:      SOFTWARE: Patent Release #1.0, Version #1.30
:      CURRENT APPLICATION DATA:
:      APPLICATION NUMBER: US/09/105,678A
:      FILING DATE: 26-JUN-1998
:      PRIORITY APPLICATION DATA:
:      APPLICATION NUMBER: JP 172118/1997
:      FILING DATE: 27-JUN-1997
:      ATTORNEY/AGENT INFORMATION:
:      NAME: Conlin, David G.
:      REGISTRATION NUMBER: 27,026
:      REFERENCE/DOCKET NUMBER: 48466-342
:      TELECOMMUNICATION INFORMATION:
:      TELEPHONE: 617-523-3400
:      TELEFAX: 617-523-6440
:      INFORMATION FOR SEQ ID NO: 8:
:      SEQUENCE CHARACTERISTICS:
:      LENGTH: 31 amino acids
:      TYPE: amino acid
:      STRANDEDNESS:
:      TOPOLOGY: linear
:      MOLECULE TYPE: peptide
:      US-09-105-678A-8

```

```

Query Match          95.4%; Score 104; DB 3; Length 31;
Best Local Similarity 90.0%; Pred. No. 3.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 1 TPDINPAMYXXRGIRPVGRF 20
   ||||||| |||||||
Db 12 TPDINPAMYTGIRPVGRF 31

```

Search completed: April 17, 2001, 15:39:49
 Job time: 318 sec

THIS PAGE BLANK (USPTO)

Tue Apr 17 15:46:22 2001

us-09-446-543a-73.rag

1-15 Affiliations

137

File GSKY
Page 1 of 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:38:42 ; Search time 116.94 Seconds
(without alignments)
10.754 Million cell updates/sec

Title: US-09-446-543A-73

Perfect score: 109

Sequence: 1 TPDIINPAWYXXRGIRPGVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 100%

Listing first 45 summaries

Database : A.GeneSeq_0401:*

1: /SIDSI/gcgdata/geneSeq/AA1980.DAT:*

2: /SIDSI/gcgdata/geneSeq/AA1981.DAT:*

3: /SIDSI/gcgdata/geneSeq/AA1982.DAT:*

4: /SIDSI/gcgdata/geneSeq/AA1983.DAT:*

5: /SIDSI/gcgdata/geneSeq/AA1984.DAT:*

6: /SIDSI/gcgdata/geneSeq/AA1985.DAT:*

7: /SIDSI/gcgdata/geneSeq/AA1986.DAT:*

8: /SIDSI/gcgdata/geneSeq/AA1987.DAT:*

9: /SIDSI/gcgdata/geneSeq/AA1988.DAT:*

10: /SIDSI/gcgdata/geneSeq/AA1989.DAT:*

11: /SIDSI/gcgdata/geneSeq/AA1990.DAT:*

12: /SIDSI/gcgdata/geneSeq/AA1991.DAT:*

13: /SIDSI/gcgdata/geneSeq/AA1992.DAT:*

14: /SIDSI/gcgdata/geneSeq/AA1993.DAT:*

15: /SIDSI/gcgdata/geneSeq/AA1994.DAT:*

16: /SIDSI/gcgdata/geneSeq/AA1995.DAT:*

17: /SIDSI/gcgdata/geneSeq/AA1996.DAT:*

18: /SIDSI/gcgdata/geneSeq/AA1997.DAT:*

19: /SIDSI/gcgdata/geneSeq/AA1998.DAT:*

20: /SIDSI/gcgdata/geneSeq/AA1999.DAT:*

21: /SIDSI/gcgdata/geneSeq/AA2000.DAT:*

22: /SIDSI/gcgdata/geneSeq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	96.3	20	18	Human type G prote
2	105	96.3	20	20	Human type G prote
3	105	96.3	20	21	Human type G prote
4	105	96.3	20	21	Human type G prote
5	105	96.3	21	18	Human type G prote
6	105	96.3	21	21	Human type G prote
7	105	96.3	22	18	Human type G prote
8	105	96.3	22	21	Human type G prote
9	105	96.3	31	18	Human type G prote
10	105	96.3	31	20	Human type G prote
11	105	96.3	31	20	Human type G prote

Result ID	Score	Query Match	Length	ID	Description
12	105	96.3	31	21	Human type G prote
13	105	96.3	31	21	Human type G prote
14	105	96.3	32	18	Human type G prote
15	105	96.3	32	21	Human type G prote
16	105	96.3	33	18	Human type G prote
17	105	96.3	33	21	Human type G prote
18	105	96.3	37	20	Human type G prote
19	105	96.3	87	20	Human type G prote
20	105	96.3	87	21	Human type G prote
21	104	95.4	20	18	Human type G prote
22	104	95.4	20	18	Human type G prote
23	104	95.4	20	20	Human type G prote
24	104	95.4	20	20	Human type G prote
25	104	95.4	20	20	Human type G prote
26	104	95.4	20	20	Human type G prote
27	104	95.4	20	21	Human type G prote
28	104	95.4	20	21	Human type G prote
29	104	95.4	20	21	Human type G prote
30	104	95.4	20	21	Human type G prote
31	104	95.4	21	18	Human type G prote
32	104	95.4	21	18	Human type G prote
33	104	95.4	21	20	Human type G prote
34	104	95.4	21	21	Human type G prote
35	104	95.4	21	21	Human type G prote
36	104	95.4	22	18	Human type G prote
37	104	95.4	22	18	Human type G prote
38	104	95.4	22	20	Human type G prote
39	104	95.4	22	21	Human type G prote
40	104	95.4	22	21	Human type G prote
41	104	95.4	31	18	Human type G prote
42	104	95.4	31	18	Human type G prote
43	104	95.4	31	20	Human type G prote
44	104	95.4	31	20	Human type G prote
45	104	95.4	31	21	Human type G prote

ALIGNMENTS

Result ID	Score	Query Match	Length	ID	Description
1	105	96.3	20	18	Human type G prote
2	105	96.3	20	20	Human type G prote
3	105	96.3	20	21	Human type G prote
4	105	96.3	20	21	Human type G prote
5	105	96.3	21	18	Human type G prote
6	105	96.3	21	21	Human type G prote
7	105	96.3	22	18	Human type G prote
8	105	96.3	22	21	Human type G prote
9	105	96.3	31	18	Human type G prote
10	105	96.3	31	20	Human type G prote
11	105	96.3	31	20	Human type G prote

XX Ligand peptide for G protein-coupled receptor - acts by modulating
PT function in the central nervous system, pancreas and pituitary gland
XX
XX Claim 2, Page 185, 258pp: English.
CC This sequence represents a peptide fragment from a novel human type
CC ligand polypeptide corresponding to amino acid residues 34 to 53 of the
CC sequence represented in W31390 and is used in an assay to monitor ligand
CC binding to the G protein-coupled receptor protein. Pharmaceutical
CC compositions containing this ligand may be used as a pituitary function
CC modulator, a central nervous system modulator or a pancreatic function
CC modulator. This ligand could have specific applications as a prophylactic
CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidemia,
CC hypercholesterolemia, hyperglycemia, hyperprolactinemia, diabetes,
CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,
CC amyotrophic lateral sclerosis, acute myocardial infarction, infertility,
CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,
CC osteoporosis and/or oligogalactia. Assays can also be developed to screen
CC compounds which are capable of altering the binding activity of the
CC ligand affecting activation of the G protein-coupled receptor protein.
SQ Sequence 20 AA:

Query Match 96.3%; Score 105; DB 18; Length 20;
Best Local Similarity 90.0%; Pred. No. 1.3e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
||||||| |||||||
DB 1 tpdinpawyasrgirpvgrf 20

RESULT 2
W97236
ID W97236 standard; peptide: 20 AA.
XX
AC W97236;
XX
DT 06-MAY-1999 (first entry)
XX
DE Human type ligand polypeptide fragment.
XX
XX Rat type ligand; modulation; prolactin secretion;
KW G protein-coupled receptor; GPCR; hypovarianism; gonecyst cacogenesis;
KW menopausal syndrome; eutthyroid; hypometabolism; lactation;
KW pituitary adenomatosis; brain tumour; emmenopathy; autoimmune disease;
KW prolactinoma; infertility; impotence; amenorrhea; galactorrhea;
KW acromegaly; Chiari-Frommel syndrome; Argonz-del Castillo syndrome;
KW Forbes-Albright syndrome; lymphoma; Sheehan syndrome; dyszoospermia;
KW contraceptive; placental function; choriocarcinoma; hydatid mole;
KW interruption mole; abortion; unfertilized fetus; abnormal saccharometabolism;
KW abnormal lipidmetabolism; oxytocia.
XX
OS Homo sapiens.
XX
PN WO9858962-A1.
XX
PD 30-DEC-1998.
XX
PF 22-JUN-1998; 98WO-JP02765.
XX
PR 23-JUN-1997; 97JP-0165437.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
XX
DR WPI: 1999-105614/09.

XX Use of G protein-coupled receptor ligands - for modulating prolactin
PT secretion or placental function, e.g. for treating menopausal
PT syndrome, tumours, autoimmune disease or abnormal pregnancy
XX
XX Claim 3, Page 166; 241pp: English.
CC The present sequence represents a human type ligand fragment. It
CC is used in the course of the invention. The specification describes
CC an agent for modulating prolactin secretion which comprises a
CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
CC protein. The agents for promoting prolactin secretion can be used for
CC treating or preventing hypovarianism, gonecyst cacogenesis, menopausal
CC syndrome, eutthyroid or hypometabolism. They can be used for promoting
CC lactation in a domestic mammal and as an aphrodisiac. The agents for
CC inhibiting prolactin secretion can be used for treating or preventing
CC pituitary adenomatosis, brain tumour, emmenopathy, autoimmune disease,
CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
CC acromegaly, Chiari-Frommel syndrome, Argonz-del Castillo syndrome,
CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dyszoospermia.
CC The inhibitory agents can also be used as contraceptives. The agents for
CC modulating placental function can be used for treating or preventing
CC choriocarcinoma, hydatid mole, abortion, unfertilized fetus,
CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
SQ Sequence 20 AA:

Query Match 96.3%; Score 105; DB 20; Length 20;
Best Local Similarity 90.0%; Pred. No. 1.3e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
||||||| |||||||
DB 1 tpdinpawyasrgirpvgrf 20

RESULT 3
B10365
ID B10365 standard; peptide: 20 AA.
XX
AC B10365;
XX
DT 24-NOV-2000 (first entry)
XX
DE Human oxytocin secretion promoting peptide SEQ ID NO: 35.
XX
XX Human; oxytocin secretion promoter; G protein-coupled receptor protein;
KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
KW caesarean section; artificial fertilization; galactostasis; goat; pig;
KW veterinary medicine; milk production.
XX
OS Homo sapiens.
XX
PN WO200038704-A1.
XX
PD 06-JUL-2000.
XX
PF 22-DEC-1999; 99WO-JP07199.
XX
PR 25-DEC-1998; 98JP-0369585.
XX
PA (TAKE) TAKEDA CHEM IND LTD.
XX
PI Matsumoto H, Kitada C, Hinuma S;
XX
DR WPI: 2000-452298/39.
XX
PT Physiologically-active polypeptide recognized as ligand by G
PT protein-coupled receptor protein, for promoting secretion of oxytocin,
PT as drugs for diseases relating to oxytocin secretion and in veterinary
PT medicine -
XX

Query Match 96.3%; Score 105; DB 18; Length 21;
Best Local Similarity 90.0%; Pred. No. 1.4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMWXXRGIRPVGRF 20
|||||
Db 1 tpdinpawyasrgirpvgrf 20

RESULT 6

B10366 B10366 standard; peptide: 21 AA.

AC B10366;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 36.

KM Human; oxytocin secretion promoter; G protein-coupled receptor protein;

KM treatment; disease: pain; atonic bleeding; uterine recovery failure; cow;

KM caesarean section; artificial fertilization; galactostasis; goat; pig;

KM veterinary medicine; milk production.

OS Homo sapiens.

PN WO200038704-A1.

PD 06-JUL-2000.

PF 22-DEC-1999; 99WO-JP07199.

PR 25-DEC-1998; 98JP-0369585.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Matsumoto H, Kitada C, Hinuma S;

DR WPI: 2000-452298/39.

PT Physiologically-active polypeptide recognized as ligand by G

PT protein-coupled receptor protein, for promoting secretion of oxytocin,

PT as drugs for diseases relating to oxytocin secretion and in veterinary

PT medicine -

PS Disclosure: Page 63; 72pp; Japanese.

CC This invention describes a novel oxytocin secretion-regulating agent

CC which contains a ligand peptide or its salt for the G protein-coupled

CC receptor protein. It is useful in the form of drugs for ameliorating,

CC preventing and treating diseases relating to oxytocin secretion e.g.

CC weak pains and atonic bleeding, before and after expulsion of placenta,

CC uterine recovery failure, caesarean section, stoppage of artificial

CC fertilization or galactostasis and is also applicable in veterinary

CC medicine for promoting milk production in cow, goat and pig. This

CC sequence represents a human peptide which acts as an oxytocin secretion

CC promoter.

SO Sequence 21 AA:

Query Match 96.3%; Score 105; DB 21; Length 21;
Best Local Similarity 90.0%; Pred. No. 1.4e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMWXXRGIRPVGRF 20
|||||
Db 1 tpdinpawyasrgirpvgrf 20

RESULT 7

B10367 B10367 standard; peptide: 22 AA.

ID W31396 standard; Peptide: 22 AA.

AC W31396;

DT 06-APR-1998 (first entry)

DE Human type G protein-coupled receptor ligand fragment 6.

KM G protein-coupled receptor; ligand binding; pharmaceutical;

KM modulator; pituitary; central nervous system; pancreas; prophylactic;

KM therapeutic agent.

OS Homo sapiens.

PN WO9724436-A2.

PD 10-JUL-1997.

PF 26-DEC-1996; 96WO-JP03821.

PR 18-SEP-1996; 96JP-0246573.

PR 28-DEC-1995; 95JP-0343371.

PR 15-MAR-1996; 96JP-0059419.

PR 12-AUG-1996; 96JP-0211805.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;

PI Kawamata Y, Kitada C;

DR WPI: 1997-363672/33.

DR N-PSDB: V02433.

PT Ligand peptide for G protein-coupled receptor - acts by modulating

PT function in the central nervous system, pancreas and pituitary gland

PS Claim 2; Page 186; 258pp; English.

CC This sequence represents a peptide fragment from a novel human type

CC ligand polypeptide corresponding to amino acid residues 34 to 55 of the

CC sequence represented in W31390 and is used in an assay to monitor ligand

CC binding to the G protein-coupled receptor protein. Pharmaceutical

CC compositions containing this ligand may be used as a pituitary function

CC modulator, a central nervous system modulator or a pancreatic function

CC modulator. This ligand could have specific applications as a prophylactic

CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,

CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,

CC growth hormone secretory disease, hyper- and hypoparathyroidism, hyperlipidaemia,

CC hypercholesterolaemia, hyperglycaemia, hyperprolactinaemia, diabetes,

CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,

CC rheumatoid arthritis, spinal injury, transient brain ischaemia, epilepsy,

CC amyotrophic lateral sclerosis, acute myocardial infarction, infertility,

CC spinocerebellar degeneration, bone fracture, trauma, atopic dermatitis,

CC osteoporosis and/or oligosaccharia. Assays can also be developed to screen

CC compounds which are capable of altering the binding activity of the

CC ligand affecting activation of the G protein-coupled receptor protein.

SO Sequence 22 AA:

Query Match 96.3%; Score 105; DB 18; Length 22;
Best Local Similarity 90.0%; Pred. No. 1.5e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMWXXRGIRPVGRF 20
|||||
Db 1 tpdinpawyasrgirpvgrf 20

RESULT 8

B10367 B10367 standard; peptide: 22 AA.

KW Irruption mole: abortion; unthrifty fetus; abnormal saccharometabolism;
 KW abnormal lipidmetabolism; oxytocia.
 XX Homo sapiens.
 OS W09858962-A1.
 PN
 PD 30-DEC-1998.
 XX
 PF 22-JUN-1998; 98WO-JP02765.
 XX
 PR 23-JUN-1997; 97JP-0165437.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Fujii R, Hinuma S, Kawamata Y, Matsumoto H;
 DR WPI: 1999-105614/09.
 XX
 PT Use of G protein-coupled receptor ligands - for modulating prolactin
 PT secretion or placental function, e.g. for treating menopausal
 PT syndrome, tumors, autoimmune disease or abnormal pregnancy
 XX
 XX Claim 3: Page 159; 241pp: English.
 PS
 XX The present sequence represents a human type ligand fragment. It
 CC is used in the course of the invention. The specification describes
 CC an agent for modulating prolactin secretion which comprises a
 CC ligand polypeptide or a salt, for a G protein-coupled receptor (GPCR)
 CC protein. The agents for promoting prolactin secretion can be used for
 CC treating or preventing hypovarianism, gonocyst cacogenesis, menopausal
 CC syndrome, euthyroid or hypometabolism. They can be used for promoting
 CC lactation in a domestic mammal and as an aphrodisiac. The agents for
 CC inhibiting prolactin secretion can be used for treating or preventing
 CC pituitary adenomatosis, brain tumor, amenorrhoea, galactorrhea,
 CC prolactinoma, infertility, impotence, amenorrhea, galactorrhea,
 CC acromegaly, Chiari-Frommel syndrome, Argon-del Castillo syndrome,
 CC Forbes-Albright syndrome, lymphoma, Sheehan syndrome or dysospermia.
 CC The inhibitory agents can also be used as contraceptives. The agents for
 CC modulating placental function can be used for treating or preventing
 CC chorioncarcinoma, hydatid mole, abortion, unthrifty fetus,
 CC abnormal saccharometabolism, abnormal lipidmetabolism or oxytocia.
 CC
 XX Sequence 31 AA:
 SQ

Query Match 96.3%; Score 105; DB 20; Length 31;
 Best Local Similarity 90.0%; Pred. No. 2.1e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
 |||||
 DB 12 tpdinpamyasrgirpvgrf 31

RESULT 11
 W87615 standard; Peptide: 31 AA.
 ID W87615:
 AC W87615:
 XX
 DT 29-MAR-1999 (first entry)
 XX
 DE Human 19P2 ligand.
 XX
 KW 19P2 ligand; G protein coupled receptor; pituitary;
 KW prolactin releasing peptide; human; dementia; breast cancer;
 KW therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP887317-A2.
 XX

PD 30-DEC-1998.
 XX
 PF 25-JUN-1998; 98EP-0111725.
 XX
 PR 27-JUN-1997; 97JP-0172118.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Moriya T, Nishimura O, Suenaga M, Tanaka Y;
 DR WPI: 1999-047884/05.
 XX
 PT Producing a 19P2 pituitary G protein receptor ligand - by cleavage
 PT of a fusion protein, useful for preventing and treating dementia,
 PT breast cancer, renal failure and autoimmune disease
 XX
 XX Claim 5; Page 35; 56pp: English.
 PS
 XX This is the amino acid sequence of the human pituitary G
 CC protein-coupled receptor ligand 19P2L. A method suitable for
 CC commercial high-level production of 19P2L comprises expressing
 CC the ligand in host cells as a recombinant fusion protein e.g. with
 CC human basic fibroblast growth factor (see V83796-97) that has
 CC been modified to include an N-terminal cysteine residue. The
 CC ligand is released from the fusion by cyanidation followed by
 CC ammonolysis. 19P2L has prolactin secretion-stimulating and (at
 CC high doses) prolactin secretion-inhibiting properties. It can be
 CC used in the treatment and prevention of various diseases including:
 CC senile dementia, cerebrovascular dementia, and dementia associated
 CC with: genealogical disorders (e.g. Alzheimer's disease, Parkinson's
 CC disease, Pick's disease, Huntington's disease), infectious diseases
 CC (e.g. Creutzfeldt-Jakob's), endocrine or metabolic disease or
 CC toxicosis (e.g. hypothyroidism, vitamin B12 deficiency, alcoholism,
 CC intoxication by drugs, metal and organic compounds), tumorigenic
 CC diseases (e.g. brain tumour), traumatic diseases (e.g. chronic
 CC subarachnoidal haemorrhage, and other types of dementia, depression,
 CC hyperactive child syndrome (microencephalopathy) and disturbance of
 CC consciousness. It is also useful for prevention and treatment of
 CC diseases associated with prolactin hypo and hypersecretion
 CC respectively, including: hyperprolactinaemia, pituitary adenoma,
 CC breast cancer, infertility, impotence and autoimmune disease
 CC (hypersecretion disorders), and seminal vesicle hypoplasia,
 CC osteoporosis, menopausal syndrome and renal failure (hyposecretion
 CC disorders). The 19P2 polypeptide/amide is also useful as a test
 CC reagent for study of the prolactin secretory function or as a
 CC laccogogue in mammalian farm animals.
 CC
 XX Sequence 31 AA:
 SQ

Query Match 96.3%; Score 105; DB 20; Length 31;
 Best Local Similarity 90.0%; Pred. No. 2.1e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
 |||||
 DB 12 tpdinpamyasrgirpvgrf 31

RESULT 12
 B10362 standard; peptide: 31 AA.
 ID B10362:
 AC B10362:
 XX
 DT 24-NOV-2000 (first entry)
 XX
 DE Human oxytocin secretion promoting peptide SEQ ID NO: 32.
 XX
 KW Human; oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease; pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.
 KW

XX OS Homo sapiens.
XX PN WO200038704-A1.
XX PD 06-JUL-2000.
XX PF 22-DEC-1999; 99WO-JP07199.
XX PR 25-DEC-1998; 98JP-0369585.
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX PI Matsumoto H, Kitada C, Hinuma S;
XX DR WPI: 2000-452298/39.
XX PT Physiologically-active polypeptide recognized as ligand by G
XX PR protein-coupled receptor protein, for promoting secretion of oxytocin,
XX PT as drugs for diseases relating to oxytocin secretion and in veterinary
XX PT medicine -
XX PS Disclosure: Page 62; 72pp; Japanese.
XX CC This invention describes a novel oxytocin secretion-regulating agent
XX CC which contains a ligand peptide or its salt for the G protein-coupled
XX CC receptor protein. It is useful in the form of drugs for ameliorating,
XX CC preventing and treating diseases relating to oxytocin secretion e.g.,
XX CC wear pains and atonic bleeding, before and after expulsion of placenta,
XX CC uterine recovery failure, caesarean section, stoppage of artificial
XX CC fertilization or galactostasis and is also applicable in veterinary
XX CC medicine for promoting milk production in cow, goat and pig. This
XX CC sequence represents a human peptide which acts as an oxytocin secretion
XX CC promoter.
XX SQ Sequence 31 AA:
SQ
Query Match 96.3%; Score 105; DB 21; Length 31;
Best Local Similarity 90.0%; Pred. No. 2.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXXRGIRPVGRF 20
DB 12 tpdinpawysrgirpvgrf 31
RESULT 13
Y49291
ID Y49291 standard; peptide; 31 AA.
XX AC Y49291;
XX DT 22-FEB-2000 (first entry)
XX DE 19P2 ligand peptide fragment.
XX KM Monoclonal antibody; 19P2 ligand; diagnosis; prolactin secretion;
XX KM pituitary; regulatory mechanism; central nervous system; pancreatic.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Modified-site 31 /note= "C-terminal amide"
XX PN WO9960112-A1.
XX PD 25-NOV-1999.
XX PF 20-MAY-1999; 99WO-JP02650.
XX PR 21-MAY-1998; 98JP-0140293.

XX XX (TAKE) TAKEDA CHEM IND LTD.
XX PA Matsumoto H, Kitada C, Hinuma S;
XX PI WPI: 2000-039381/03.
XX DR
XX PT New monoclonal antibodies, useful in diagnosis, as drugs and in
XX PT studying diseases related to ligand abnormality -
XX PS Disclosure: Page 26; 73pp; Japanese.
XX CC The invention provides a monoclonal antibody which has a specific
XX CC reaction with the part peptide of the C-terminal of 19P2 ligand or its
XX CC derivative. The antibodies can be used in diagnosis or to treat or
XX CC prevent diseases associated with abnormality in the pituitary function
XX CC regulatory mechanism (e.g. promotion of prolactin secretion), central
XX CC nervous regulatory mechanism, and pancreatic function regulatory
XX CC mechanism. The antibody-based immunoassay can also be applied in
XX CC clarifying the physiological functions of the ligand and its derivative.
XX CC Sequences Y49290-302 represent peptide fragments of the 19P2 ligand.
XX SQ Sequence 31 AA:
SQ
Query Match 96.3%; Score 105; DB 21; Length 31;
Best Local Similarity 90.0%; Pred. No. 2.1e-11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 1 TPDINPAMYXXRGIRPVGRF 20
DB 12 tpdinpawysrgirpvgrf 31

RESULT 14
W31392
ID W31392 standard; Peptide; 32 AA.
XX AC W31392;
XX DT 06-APR-1998 (first entry)
XX DE Human type G protein-coupled receptor ligand fragment 2.
XX KM G protein-coupled receptor; ligand binding; pharmaceutical;
XX KM modulator; pituitary; central nervous system; pancreas; prophylactic;
XX KM therapeutic agent.
XX OS Homo sapiens.
XX PN WO9724436-A2.
XX PD 10-JUL-1997.
XX PF 26-DEC-1996; 96WO-JP03821.
XX PR 18-SEP-1996; 96JP-0246573.
XX PR 28-DEC-1995; 95JP-0343371.
XX PR 15-MAR-1996; 96JP-0059419.
XX PR 12-AUG-1996; 96JP-0211805.
XX PA (TAKE) TAKEDA CHEM IND LTD.
XX PI Fujii R, Fukusumi S, Habata Y, Hinuma S, Hosoya M;
XX PI Kawamata Y, Kitada C;
XX DR WPI: 1997-363672/33.
XX DR N-PSDB; V02429.
XX PT Ligand peptide for G protein-coupled receptor - acts by modulating
XX PT function in the central nervous system, pancreas and pituitary gland
XX SQ Claim 2; Page 185; 258pp; English.

XX This sequence represents a peptide fragment from a novel human type
 CC ligand polypeptide corresponding to amino acid residues 23 to 54 of the
 CC sequence represented in W31390 and is used in an assay to monitor ligand
 CC binding to the G protein-coupled receptor protein. Pharmaceutical
 CC compositions containing this ligand may be used as a pituitary function
 CC modulator, a central nervous system modulator or a pancreatic function
 CC modulator. This ligand could have specific applications as a prophylactic
 CC or therapeutic agent for dementia, depression, hyperkinetic syndrome,
 CC disturbance of consciousness, anxiety syndrome, schizophrenia, trauma,
 CC growth hormone secretory disease, hyper- and polyphagia, hyperlipidaemia,
 CC hypercholesterolaemia, hyperglycaemia, hyperprolactinaemia, diabetes,
 CC cancer, pancreatitis, renal disease, Turner's syndrome, neurosis, asthma,
 CC rheumatoid arthritis, spinal injury, transient brain ischemia, epilepsy,
 CC amyotrophic lateral sclerosis, acute myocardial infarction, infertility,
 CC splintered lateral degeneration, bone fracture, trauma, atopic dermatitis,
 CC osteoporosis and/or oligosaccharia. Assays can also be developed to screen
 CC compounds which are capable of altering the binding activity of the
 CC ligand affecting activation of the G protein-coupled receptor protein.

XX SQ Sequence 32 AA;

Query Match 96.3%; Score 105; DB 18; Length 32;
 Best Local Similarity 90.0%; Pred. No. 2.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
 |||||
 DB 12 tpdlnpawysrglrpvgrf 31

RESULT 15

ID B10363 standard; peptide: 32 AA.

XX B10363;

DT 24-NOV-2000 (first entry)

DE Human oxytocin secretion promoting peptide SEQ ID NO: 33.

XX Human: oxytocin secretion promoter; G protein-coupled receptor protein;
 KW treatment; disease: pain; atonic bleeding; uterine recovery failure; cow;
 KW caesarean section; artificial fertilization; galactostasis; goat; pig;
 KW veterinary medicine; milk production.

XX Homo sapiens.

XX WO200038704-A1.

XX 06-JUL-2000.

XX 22-DEC-1999; 99WO-JP07199.

XX 25-DEC-1998; 98JP-0369585.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Matsumoto H, Kitada C, Hinuma S;

XX WPI; 2000-452298/39.

XX Physiologically-active polypeptide recognized as ligand by G
 PT protein-coupled receptor protein, for promoting secretion of oxytocin,
 PT as drugs for diseases relating to oxytocin secretion and in veterinary
 PT medicine -

XX Disclosure; Page 62; 72pp; Japanese.

XX This invention describes a novel oxytocin secretion-regulating agent
 CC which contains a ligand peptide or its salt for the G protein-coupled
 CC receptor protein. It is useful in the form of drugs for ameliorating,

CC preventing and treating diseases relating to oxytocin secretion e.g.
 CC weak pains and atonic bleeding, before and after expulsion of placenta,
 CC uterine recovery failure, caesarean section, stoppage of artificial
 CC fertilization or galactostasis and is also applicable in veterinary
 CC medicine for promoting milk production in cow, goat and pig. This
 CC sequence represents a human peptide which acts as an oxytocin secretion
 CC promoter.

XX SQ Sequence 32 AA;

Query Match 96.3%; Score 105; DB 21; Length 32;
 Best Local Similarity 90.0%; Pred. No. 2.2e-11;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TPDINPAMYXXRGIRPVGRF 20
 |||||
 DB 12 tpdlnpawysrglrpvgrf 31

Search completed: April 17, 2001, 15:38:42
 Job time: 349 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:45:59 ; Search time 70.08 seconds

(without alignments)
21.574 Million cell updates/sec

Title: US-09-446-543a-73

Perfect score: 109

Sequence: 1 TPDINPAWYXXRGIRPVGRFX 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR.67:*

2: pir1:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49.5	45.4	664	2 F83376	conserved hypotet
2	49	45.0	790	2 T47959	hypothetical prote
3	46	42.2	333	2 H82852	hydroxybenzoate oc
4	45	41.3	767	2 T21969	hypothetical prote
5	44	40.4	250	2 G83400	hypothetical prote
6	44	40.4	284	2 F71015	hypothetical prote
7	43.5	39.9	1501	2 T45623	hypothetical prote
8	43	38.4	220	2 C83292	probable glutathio
9	43	38.4	232	2 G75608	hypothetical prote
10	43	39.4	309	2 T32376	hypothetical prote
11	43	39.4	376	2 A48197	opsin, ocellar - A
12	43	39.4	376	2 B48197	opsin, lateral eye
13	43	39.4	476	2 G64720	probable amino aci
14	43	38.4	719	2 S61046	ARPI protein - yea
15	43	38.4	1296	2 T16859	hypothetical prote
16	42.5	39.0	443	2 T21499	hypothetical prote
17	42	38.5	428	2 F81694	pyruvate dehydroge
18	41.5	38.1	345	2 D84012	N-acetylglutamate
19	41	37.6	128	2 S76955	hypothetical prote
20	41	37.6	284	2 A75117	hypothetical prote
21	41	37.6	338	2 T20100	hypothetical prote
22	41	37.6	342	2 B64395	malic acid transpo
23	41	37.6	343	2 T46534	malic acid transpo
24	41	37.6	347	2 H64371	malic acid transpo
25	41	37.6	986	1 OYURGA	spectrin receptor p
26	41	37.6	2155	2 T30197	alpha tectorin - m
27	40.5	37.2	779	2 T49717	related to BCS1 pr
28	40.5	37.2	1540	2 T45619	hypothetical prote
29	40	36.7	184	2 T35841	probable membrane

30	40	36.7	324	2 T35901	probable arac fam1
31	40	36.7	329	2 H70744	hypothetical prote
32	40	36.7	341	2 T35426	probable oxidoredu
33	40	36.7	359	2 T40084	PMP domain protei
34	40	36.7	390	2 G82844	cysteine synthase
35	40	36.7	430	1 B69009	conserved hypotet
36	40	36.7	462	2 T00708	violaxanthin de-ep
37	40	36.7	468	2 C83160	nitrite extrusion
38	40	36.7	546	2 A33260	cholesterol oxidas
39	40	36.7	581	2 A42743	pol polyprotein -
40	40	36.7	843	1 GNVWK	pol polyprotein -
41	40	36.7	1196	1 GNMVGV	pol polyprotein -
42	40	36.7	1196	1 GNMVGV	pol polyprotein -
43	40	36.7	1762	2 T03222	probable polykeid
44	40	36.7	2100	2 T03223	probable polykeid
45	40	36.7	7962	2 I38346	elastic titin - hu

ALIGNMENTS

RESULT 1
F83376
conserved hypothetical protein PA2151 [Imported] - Pseudomonas aeruginosa (strain PAO
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83376
R:Stover, C.K.; Pham, X.Q.; Ertwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L
.; Lofy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: F83376
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-664 <STO>
A:Cross-references: GB:AE004642; GB:AE004091; NID:g9948163; PIDN:AA05539.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2151

Query Match 45.4% Score 49.5; DB 2; Length 664;
Best Local Similarity 62.5%; Pred. No. 2.6;
Matches 10; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

OY 1 TPDINPAWYXXRGIRP 16
||||| | | | |
Db 478 TPDINP-WFLQGRGP 492

RESULT 2
T47959
hypothetical protein F15616.60 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47959
R:De Haan, M.; Maarse, A.C.; Griwell, L.A.; Newes, H.W.; Lemcke, K.; Mayer, K.F.X.; Q
submitted to the Protein Sequence Database, January 2000
A:Reference number: Z24480
A:Accession: T47959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-790 <DEH>
A:Cross-references: EMBL:AL132959
A:Experimental source: cultivar Columbia; BAC clone F15616
C:Genetics:
A:Map position: 3
A:introns: 39/1; 678/2; 698/3; 773/2
A:Note: F15616.60

```

Query Match          45.0%: Score 49; DB 2; Length 790;
Best Local Similarity 47.4%: Pred. No. 3.7;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMVXXRGIRPVGR 20
    | | | | | | | | | |
Db 366 PPHNPRTYGRGLQPHGRW 384

RESULT 3
H82852
hydroxybenzoate octaprenyltransferase XF0068 [imported] - Xylella fastidiosa (strain 9a5
C:Species: Xylella fastidiosa
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C:Accession: H82852
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: H82852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <SIM>
A:Cross-references: GB:AE003860; GB:AE003849; NID:g9104830; PIDN:AAF82881.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Britones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.U.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
chado, M.A.; Madeira, A.M.B.N.; Madeira, J.M.F.; Martino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J.; de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0068
C:Superfamily: 4-hydroxybenzoate octaprenyltransferase

Query Match          42.2%: Score 46; DB 2; Length 333;
Best Local Similarity 53.3%: Pred. No. 4.9;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 INPAMVXXRGIRPVGR 18
    | | | | | | | | | |
Db 54 LDPMWKLARGDRPVGR 68

RESULT 4
T21969
hypothetical protein F38E11.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21969
R:Matthews, P.
submitted to the EMBL Data Library, January 1996
A:Reference number: Z19495
A:Accession: T21969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-767 <MIL>
A:Cross-references: PIDN:CAAG3775.1; GSPDB:GN00022; CESP:F38E11.7
A:Experimental source: clone F38E11
C:Genetics:
A:Gene: CESP:F38E11.7
A:Map position: 4
A:Introns: 50/2; 118/1; 139/2; 189/3; 226/1; 248/1; 287/2; 375/2; 432/3; 465/3; 548/1; 6

```

```

Query Match          41.3%: Score 45; DB 2; Length 767;
Best Local Similarity 75.0%: Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 PDINPAMV 9
    | | | | |
Db 747 PDVKKPAMV 754

RESULT 5
G83400
hypothetical protein PA1952 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83400
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: G83400
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-250 <STO>
A:Cross-references: GB:AE004622; GB:AE004091; NID:g9947948; PIDN:AA05340.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1952

Query Match          40.4%: Score 44; DB 2; Length 250;
Best Local Similarity 47.1%: Pred. No. 7.8;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMVXXRGIRPVGR 18
    | | | | | | | | | |
Db 233 PALNCAMEQLRALRPSG 249

RESULT 6
F71015
hypothetical protein PH1420 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: F71015
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: F71015
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-284 <KAW>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAA30526.1; PID:g3257843
A:Experimental source: strain OT3
A>Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH1420
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1420

Query Match          40.4%: Score 44; DB 2; Length 284;
Best Local Similarity 44.4%: Pred. No. 8.9;
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMVXXRGIRPVGR 19
    | | | | | | | | | |
Db 217 PYIEPFFALRGLELLGR 234

```

```

RESULT 7
T45623
hypothetical protein F13624.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45623
R:Bevan, M.; Van Der Schueren, J.; Chuang, Y.J.; Voet, M.; Robben, J.; Volckaert, G.; Be
submitted to the Protein Sequence Database, December 1999
A:Reference number: 223009
A:Accession: T45623
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1501 <BEV>
A:Cross-references: EMBL:AL133421
A:Experimental source: cultivar Columbia; BAC clone F13624
C:Genetics:
A:Map position: 5
A:introns: 64/1; 739/3; 785/2; 1302/2; 1318/3; 1399/2; 1434/2
A>Note: F13624.180

Query Match 39.4%; Score 43.5; DB 2; Length 1501;
Best Local Similarity 34.5%; Pred. No. 57;
Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 2;

QY 1 TPDIHPAWYXXRG-----IRPV---GRF 20
DB 1184 SPQMAPSWISQYGTTRKNGLQVPVNDTRGF 1212

RESULT 8
C83292
probable glutathione S-transferase PA2821 [imported] - Pseudomonas aeruginosa (strain PA
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83292
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lm,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: C83292
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-220 <STO>
A:Cross-references: GB:AE004709; GB:AE004091; NID:g9948904; PIDN:AAG06209.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2821

Query Match 39.4%; Score 43; DB 2; Length 220;
Best Local Similarity 64.3%; Pred. No. 10;
Matches 9; Conservative 1; Mismatches 2; Indels 2; Gaps 1;

QY 6 PAMYXXRGIRPYGR 19
DB 39 PAMY--REISPIGR 50

RESULT 9
G75608
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: G75608
R:White, O.; Eelsen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vanatvean, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zaleski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

```

```

A:Reference number: A75250; MUID:20036896
A:Accession: G75608
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-232 <WHI>
A:Cross-references: GB:AE001862; GB:AE001825; NID:g6460468; PIDN:AAF12317.1; PID:g646
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR0132
A:Map position: 2

Query Match 39.4%; Score 43; DB 2; Length 232;
Best Local Similarity 46.7%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 PDINPAMYXXRGIRPYGR 16
DB 130 PDHRAWHLLRGVLP 144

RESULT 10
T32376
hypothetical protein K10F12.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32376
R:Wohlmann, P.; Beck, C.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid K10F12.
A:Reference number: 221157
A:Accession: T32376
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-309 <WOH>
A:Cross-references: EMBL:AF025462; PIDN:AAH71002.1; GSPDB:GN00021; CESP:K10F12.4
A:Experimental source: strain Bristol N2; clone K10F12
C:Genetics:
A:Gene: CESP:K10F12.4
A:Map position: 3
A:introns: 31/3; 123/2; 196/3; 239/1

Query Match 39.4%; Score 43; DB 2; Length 309;
Best Local Similarity 44.4%; Pred. No. 14;
Matches 8; Conservative 3; Mismatches 5; Indels 2; Gaps 1;

QY 2 PDINPAMYXXRGIRPYGR 19
DB 131 PDSPPMWYLPKS--PIGR 146

RESULT 11
A48197
opsin, ocellar - Atlantic horseshoe crab
C:Species: Limulus polyphemus (Atlantic horseshoe crab)
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 13-Aug-1999
C:Accession: A48197
R:Smith, W.C.; Price, D.A.; Greenberg, R.M.; Battelle, B.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 6150-6154, 1993
A:Title: Opsins from the lateral eyes and ocelli of the horseshoe crab, Limulus poly
A:Reference number: A48197; MUID:93317641
A:Accession: A48197
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-376 <SMI>
A:Cross-references: EMBL:L03792; NID:g156644; PIDN:AAA28274.1; PID:g156645
C:Superfamily: vertebrate rhodopsin
C:Keywords: chromoprotein; G protein-coupled receptor; lipoprotein; photoreceptor; re
F:318/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 39.4%; Score 43; DB 2; Length 376;

```

Best Local Similarity 46.2%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PDINPAMYXXRGI 14

Db 40 PPMNPLMYSILGV 52

RESULT 12

B48197 opsin, lateral eye - Atlantic horseshoe crab
C:Species: Limulus polyphemus (Atlantic horseshoe crab)
C:Date: 17-Jul-1994 #sequence_revision 17-Jul-1994 #text_change 13-Aug-1999
C:Accession: B48197
R:Smith, W.C.; Price, D.A.; Greenberg, R.M.; Batteille, B.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 6150-6154, 1993
A:Title: Opsins from the lateral eyes and ocelli of the horseshoe crab, Limulus polyphemus
A:Reference number: A48197; MUID:93317641
A:Accession: B48197
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-376 <SM>
A:Cross-references: EMBL:103791; NID:g156642; PIDN:AAA28273.1; PID:g156643
C:Superfamily: vertebrate rhodopsin
C:Keywords: chromoprotein; G protein-coupled receptor; lipoprotein; photoreceptor; retina
F:318/Binding site: retinal (Lys) (covalent) #status predicted

Query Match 39.4%; Score 43; DB 2; Length 376;
Best Local Similarity 46.2%; Pred. No. 17;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 PDINPAMYXXRGI 14
Db 40 PPMNPLMYSILGV 52

RESULT 13

G64720 probable amino acid transport protein yaaJ, sodium-dependent - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 26-Aug-1999
C:Accession: G64720
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64720
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-476 <BLAT>
A:Cross-references: GB:AE000111; GB:U00096; NID:g1786181; PIDN:ACC73118.1; PID:g1786188;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yaaJ
C:Superfamily: sodium-dependent D-alanine/glycine transport protein
C:Keywords: amino acid transport; transmembrane protein
F:10/26/Domain: transmembrane #status predicted <TM1>
F:91-107/Domain: transmembrane #status predicted <TM2>
F:142-155/Domain: transmembrane #status predicted <TM3>
F:178-194/Domain: transmembrane #status predicted <TM4>
F:208-224/Domain: transmembrane #status predicted <TM5>
F:303-319/Domain: transmembrane #status predicted <TM6>
F:349-365/Domain: transmembrane #status predicted <TM7>
F:391-407/Domain: transmembrane #status predicted <TM8>
F:414-430/Domain: transmembrane #status predicted <TM9>

Query Match 39.4%; Score 43; DB 2; Length 476;
Best Local Similarity 44.4%; Pred. No. 22;
Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

QY 3 DIN-----PAMYXXRGI 14
Db 120 DVNGQFRCGPAMYMARGL 137

RESULT 14

S61046 ARPI protein - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein D1478; protein YDL167c
C:Species: Saccharomyces cerevisiae
C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
C:Accession: S61046; S31139; S67719
R:Polh, T.M.
submitted to the EMBL Data Library, November 1995
A:Reference number: S61010
A:Accession: S61046
A:Molecule type: DNA
A:Residues: 1-719 <POH>
A:Cross-references: EMBL:Z67750; NID:g1061256; PIDN:CAA91579.1; PID:g1061272
R:Wehner, E.P.; Rao, E.; Brendel, M.
Mol. Gen. Genet. 237, 351-358, 1993
A:Title: Molecular structure and genetic regulation of SFA, a gene responsible for re
A:Reference number: S31138; MUID:93247548
A:Accession: S31139
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-492, 'N', 494-719 <WEH>
A:Cross-references: EMBL:X68020; NID:g577609; PIDN:CAA48159.1; PID:g288590
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
R:Polh, T.M.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67708
A:Accession: S67719
A:Molecule type: DNA
A:Residues: 1-719 <POH>
A:Cross-references: EMBL:Z74215; NID:g1431265; PIDN:CAA98741.1; PID:e253076; PID:g143
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:NRP1
A:Cross-references: MIPS:YDL167c; SGD:S0002326
A:Map position: 4L

Query Match 39.4%; Score 43; DB 2; Length 719;
Best Local Similarity 50.0%; Pred. No. 33;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 AMYXXRGIRPVG 18
Db 244 SWFTQYGVAPVG 255

RESULT 15

T16859 hypothetical protein T13C2.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C:Accession: T16859
R:Du, Z.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid T13C2.
A:Reference number: Z18591
A:Accession: T16859
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1296 <DDZ>
A:Cross-references: EMBL:U40030; NID:g1055164; PID:g1055165; PIDN:AAA81133.1; CESP:T1
C:Genetics:
A:Gene: CESP:T13C2.5
A:Introns: 62/3; 93/3; 124/3; 142/1; 212/1; 293/1; 336/2; 406/3; 437/1; 508/1; 584/1;

Query Match 39.4%; Score 43; DB 2; Length 1296;

Best Local Similarity 45.0%; Pred. No. 60;
Matches 9; Conservative 2; Mismatches 7; Indels 2; Gaps 1;

QY 1 TPDINPAWYXXRCIRPVGRF 20

DB 975 TTDINSDWYFSR--KDINRF 992

Search completed: April 17, 2001, 15:45:59
Job time: 607 sec

THIS PAGE BLANK (USPTO)

10 9 8 7 6 5 4 3 2 1

Query Match	96.38;	Score 105;	DB 1;	Length 87;
Best Local Similarity	90.08;	Pred. No. 1.5e-10;		
Matches 18; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

```

RESULT 2
PRRP_RAT STANDARD: PRT: 83 AA.
AC P81278;
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELASING PEPTIDE PRRP20].
GN PRH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsunoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RT Nature 393:272-276(1998).
CC -1- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HG3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AB015418; BAA29026.1;
CC KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
CC FT SIGNAL 1 21
CC FT PEPTIDE 22 52 PROLACTIN-RELEASING PEPTIDE PRRP31.
CC FT PEPTIDE 33 52 PROLACTIN-RELEASING PEPTIDE PRRP20.
CC FT MOD.RES 52 52 AMIDATION (G-53 PROVIDE AMIDE GROUP).
CC FT MOD.RES 83 AA; 9215 MW; DDC75A264EEB4F29 CRC64;
CC SQ SEQUENCE

```

Query Match Score 104; DB 1; Length 83;

Best Local Similarity 90.0%; Pred. No. 2e-10; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 TPDINPAMYXXRGIRPVGRF 20
    |||||  |||||
DB 33 TPDINPAMYTGIRPVGRF 52

```

```

RESULT 3
PRRP_BOVIN STANDARD: PRT: 98 AA.
AC P81264;
DT 30-MAY-2000 (rel. 39, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DE PROLACTIN-RELEASING PEPTIDE PRECURSOR (PRRP) (PROLACTIN-RELEASING
DE HORMONE) [CONTAINS: PROLACTIN-RELEASING PEPTIDE PRRP31; PROLACTIN-
DE RELASING PEPTIDE PRRP20].
GN PRH.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-52.

```

```

RC TISSUE=Brain;
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsunoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.;
RT "A prolactin-releasing peptide in the brain.";
RT Nature 393:272-276(1998).
CC -1- FUNCTION: STIMULATES PROLACTIN (PRL) RELEASE AND REGULATES THE
CC EXPRESSION OF PROLACTIN THROUGH ITS RECEPTOR HG3. MAY STIMULATE
CC LACTOTROPHS DIRECTLY TO SECRETE PRL.
CC -1- TISSUE SPECIFICITY: MEDULLA OBLONGATA AND HYPOTHALAMUS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL; AB015417; BAA29025.1;
CC KW Hormone; Amidation; Signal; Cleavage on pair of basic residues.
CC FT SIGNAL 1 22
CC FT PEPTIDE 23 53 PROLACTIN-RELEASING PEPTIDE PRRP31.
CC FT PEPTIDE 33 53 PROLACTIN-RELEASING PEPTIDE PRRP20.
CC FT MOD.RES 53 53 AMIDATION (G-54 PROVIDE AMIDE GROUP).
CC FT MOD.RES 98 AA; 10544 MW; 08AC35A13B0FA508 CRC64;
CC SQ SEQUENCE

```

Query Match Score 104; DB 1; Length 98;

Best Local Similarity 90.0%; Pred. No. 2.4e-10;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

OY 1 TPDINPAMYXXRGIRPVGRF 20
    |||||  |||||
DB 34 TPDINPAMYTGIRPVGRF 53

```

```

RESULT 4
EX1L_HUMAN STANDARD: PRT: 676 AA.
AC Q92935;
DT 15-JUL-1998 (rel. 36, Created)
DT 15-JUL-1998 (rel. 36, Last sequence update)
DE 01-OCT-2000 (rel. 40, Last annotation update)
DE EXOSTOSIN-LIKE 1 (EXOSTOSIN-L) (MULTIPLE EXOSTOSIS-LIKE PROTEIN).
GN EX1L OR EXTL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97189339; PubMed=9037597;
RA Wise C.A., Clines G.A., Massa H., Trask B.J., Lovett M.;
RT "Identification and localization of the gene for EXTL, a third member
RT of the multiple exostosins gene family.";
RT genome Res. 7:10-16(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Xu L., Deng H.X., Xia J.H., Pan Q., Liu C.Y.;
RT "Mutations of the EXT genes in hereditary multiple exostoses in
RT Chinese.";
RT Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Wuyts W., Spieker N., Van Roy N., De Paeppe A., De Boule K.,
RA Willems P.J., Van Hul W., Versteeg R., Spelman F.;
RT "Refined physical mapping and genomic structure of the EXTL1 gene.";
RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: APPEARS TO BE A TUMOR SUPPRESSOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. ENDOPLASMIC
CC RETICULUM (BY SIMILARITY).

```

CC -1- SIMILARITY: BELONGS TO THE EXOSTOSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: U67191; AAC51141.1; -;
DR EMBL: AF083633; AAD02840.1; -;
DR EMBL: AF083623; AAD02840.1; JOINED.
DR EMBL: AF083624; AAD02840.1; JOINED.
DR EMBL: AF083625; AAD02840.1; JOINED.
DR EMBL: AF083626; AAD02840.1; JOINED.
DR EMBL: AF083627; AAD02840.1; JOINED.
DR EMBL: AF083628; AAD02840.1; JOINED.
DR EMBL: AF083629; AAD02840.1; JOINED.
DR EMBL: AF083630; AAD02840.1; JOINED.
DR EMBL: AF083631; AAD02840.1; JOINED.
DR EMBL: AF083632; AAD02840.1; JOINED.
DR EMBL: AF153980; AAF73172.1; -;
DR EMBL: AF151391; AAF73172.1; JOINED.
DR MIM: 601738; -;
KW Anti-oncogene; Multigene family; Transmembrane; Signal-anchor.
FT TRANSMEM 10 30 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT SEQUENCE 676 AA; 74673 MW; B5E006A8762E5633 CRC64;

OY 1 TPDIINPAWYXXRGIRPVGRF 20
DB 400 SPQDPFFYYLQGSRRPGRF 419

RESULT 5
OPSL_LIMPO STANDARD; PRT; 376 AA.
AC P35360; (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE LATERAL EYE OPSIN.
OS Limulus polyphemus (Atlantic horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Limulus.
OX NCBI_TaxID=6850;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lateral eye;
RX MEDLINE=93317641; PubMed=8327495;
RA Smith W.C., Price D.A., Greenberg R.M., Battelle B.-A.;
RT "Opsins from the lateral eyes and ocell of the horseshoe crab,
Limulus polyphemus.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: LATERAL EYE.
CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
CC BE PHOSPHORYLATED (BY SIMILARITY).
CC -1- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMUM AT 520 NM.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC OPSIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L03791; AAA28273.1; -;
DR EMBL: L03781; AAA02498.1; -;
DR PIR: B48197; B48197.
DR GCRDB: GCR_0584; -;
DR InterPro: IPR000276; -;
DR InterPro: IPR001391; -;
DR InterPro: IPR001760; -;
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS: PR00237; GPCRHHODOPSN.
DR PRINTS: PR00238; OPSIN.
DR PRINTS: PR00578; OPSINTRLEVE.
DR PROSITE: PS00237; G_PROTEIN_REC_P1_1; 1.
DR PROSITE: PS00262; G_PROTEIN_REC_P1_2; 1.
DR PROSITE: PS00238; OPSIN; 1.
KW Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; G-Protein coupled receptor.
FT DOMAIN 1 71
FT TRANSMEM 47 71
FT DOMAIN 72 83
FT TRANSMEM 84 108
FT DOMAIN 109 123
FT TRANSMEM 124 143
FT DOMAIN 144 162
FT TRANSMEM 163 186
FT DOMAIN 187 210
FT TRANSMEM 211 238
FT DOMAIN 239 274
FT TRANSMEM 275 298
FT DOMAIN 307 331
FT TRANSMEM 332 376
FT DOMAIN 377 421
FT DISULFID 120 197
FT BINDING 318 318
FT CARBOHYD 193 193
FT SEQUENCE 376 AA; 42139 MW; CCE401766AB06F26 CRC64;

OY 2 PDINPAWYXXRGIR 14
DB 40 PPMNPLWYSITGV 52

Query Match 39.4%; Score 43; DB 1; Length 376;
Best Local Similarity 46.2%; Pred. No. 8.5;
Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

RESULT 6
OPSL_LIMPO STANDARD; PRT; 376 AA.
AC P35361; (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE OCELLAR OPSIN.
OS Limulus polyphemus (Atlantic horseshoe crab).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Merostomata; Xiphosura;
OC Limulidae; Limulus.
OX NCBI_TaxID=6850;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Median ocell;
RX MEDLINE=93317641; PubMed=8327495;
RA Smith W.C., Price D.A., Greenberg R.M., Battelle B.-A.;
RT "Opsins from the lateral eyes and ocell of the horseshoe crab,
Limulus polyphemus.";

RL Proc. Natl. Acad. Sci. U.S.A. 90:6150-6154(1993).
 CC -1- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
 CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
 CC LINKED TO CIS-RETINAL.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: OCULAR CELLS; MEDIAN OCCELL.
 CC -1- PTM: SOME OR ALL OF THE CARBOXYL-TERMINAL SER OR THR RESIDUES MAY
 CC BE PHOSPHORYLATED (BY SIMILARITY).
 CC -1- MISCELLANEOUS: THIS OPSIN HAS AN ABSORPTION MAXIMA AT 530 NM.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC OPSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 CC DR EMBL: L03792; AAA28274.1; -;
 CC DR EMBL: L03782; AAA02499.1; -;
 CC DR PIR: A48197; A48197.
 CC GCRDB: GCR_0585; -;
 CC DR InterPro: IPR000276; -;
 CC DR InterPro: IPR001391; -;
 CC DR InterPro: IPR001760; -;
 CC DR Pfam: PF00001; 7cm1.1;
 CC DR PRINTS: PR00237; GPCRHDOPSN.
 CC DR PRINTS: PR00238; OPSIN.
 CC DR PRINTS: PR00578; OPSTRLEVE.
 CC DR PROSITE: PS00237; G-PROTEIN_RECIP_F1_1; 1.
 CC DR PROSITE: PS00262; G-PROTEIN_RECIP_F1_2; 1.
 CC DR PROSITE: PS00238; OPSIN; 1.
 CC KM Photoreceptor: Retinal protein: Transmembrane; Glycoprotein; Vision:
 CC FT Phosphorylation: G-protein coupled receptor.
 CC FT DOMAIN 1 46 EXTRACELLULAR.
 CC FT TRANSMEM 47 71 1 (POTENTIAL).
 CC FT DOMAIN 72 83 CYTOPLASMIC.
 CC FT TRANSMEM 84 108 2 (POTENTIAL).
 CC FT DOMAIN 109 123 EXTRACELLULAR.
 CC FT TRANSMEM 124 143 3 (POTENTIAL).
 CC FT DOMAIN 144 162 CYTOPLASMIC.
 CC FT TRANSMEM 163 186 4 (POTENTIAL).
 CC FT DOMAIN 187 210 EXTRACELLULAR.
 CC FT TRANSMEM 211 238 5 (POTENTIAL).
 CC FT DOMAIN 239 274 CYTOPLASMIC.
 CC FT TRANSMEM 275 298 6 (POTENTIAL).
 CC FT DOMAIN 299 306 EXTRACELLULAR.
 CC FT TRANSMEM 307 331 7 (POTENTIAL).
 CC FT DOMAIN 332 376 CYTOPLASMIC.
 CC FT DISULFID 120 197 BY SIMILARITY.
 CC FT BINDING 318 318 RETINAL CHROMOPHORE (BY SIMILARITY).
 CC FT CARBOHYD 17 17 N-LINKED (GLCNAC...) (POTENTIAL).
 CC FT CARBOHYD 193 193 N-LINKED (GLCNAC...) (POTENTIAL).
 CC SQ SEQUENCE 376 AA; 42111 MW; FA9647C40531CBF8 CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 376;
 Best Local Similarity 46.2%; Pred. No. 8.5;
 Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 OY 2 PDINPAMYXXRGI 14
 DB 40 PPMNPLWYSILGV 52
 RESULT 7
 CYCR_CHRVI STANDARD; PRT; 383 AA.
 ID CYCR_CHRVI
 AC 082947;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PHOTOSYNTHETIC REACTION CENTER CYTOCHROME C SUBUNIT PRECURSOR.
 GN PUPC.
 OS Chromatium vinosum.
 OC Bacteria; Proteobacteria; gamma subdivision; Chromatiaceae;
 OC Allochromatium.
 OX NCBI_TaxID=1049;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-D;
 RA Corson G.E., Nagashima K.V., Matsuura K., Sakurai Y., Ruwanti W.,
 RA Qin H., Allen R., Knaf D.B.;
 RT "Primary structure of genes encoding light-harvesting and reaction
 RT center proteins from Chromatium vinosum."
 RT Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THE REACTION CENTER OF PURPLE BACTERIA CONTAIN A
 CC TIGHTLY BOUND CYTOCHROME MOLECULE WHICH REDUCES THE PHOTO
 CC OXIDIZED PRIMARY ELECTRON DONOR.
 CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
 CC ANCHOR (BY SIMILARITY).
 CC -1- PTM: BINDS FOUR HEME GROUPS PER MOLECULE.
 CC -1- SIMILARITY: HIGH, WITH OTHER PHOTOSYNTHETIC REACTION CENTER
 CC CYTOCHROME C SUBUNITS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).
 CC -----
 CC DR EMBL: AB011811; BAA32742.1; -;
 CC DR InterPro: IPR000345; -;
 CC DR PROSITE: PS00190; CYTOCHROME_C_4
 CC KM Electron transport; Photosynthesis; Reaction center; Heme:
 CC KM Membrane; Lipoprotein; Duplication.
 CC FT SIGNAL 1 22 BY SIMILARITY.
 CC FT CHAIN 23 383 PHOTOSYNTHETIC REACTION CENTER CYTOCHROME
 CC C SUBUNIT.
 CC FT LIPID 23 23 N-ACYL DIGLYCERIDE (BY SIMILARITY).
 CC FT BINDING 107 107 HEME 1 (COVALENT) (BY SIMILARITY).
 CC FT BINDING 110 110 HEME 1 (COVALENT) (BY SIMILARITY).
 CC FT METAL 111 111 IRON 1 (HEME AXIAL LIGAND)
 CC (BY SIMILARITY).
 CC FT BINDING 152 152 HEME 2 (COVALENT) (BY SIMILARITY).
 CC FT BINDING 155 155 HEME 2 (COVALENT) (BY SIMILARITY).
 CC FT METAL 156 156 IRON 2 (HEME AXIAL LIGAND)
 CC (BY SIMILARITY).
 CC FT BINDING 247 247 HEME 3 (COVALENT) (BY SIMILARITY).
 CC FT BINDING 250 250 HEME 3 (COVALENT) (BY SIMILARITY).
 CC FT METAL 251 251 IRON 3 (HEME AXIAL LIGAND)
 CC (BY SIMILARITY).
 CC FT BINDING 307 307 HEME 4 (COVALENT) (BY SIMILARITY).
 CC FT BINDING 310 310 HEME 4 (COVALENT) (BY SIMILARITY).
 CC FT METAL 311 311 IRON 4 (HEME AXIAL LIGAND)
 CC (BY SIMILARITY).
 CC SQ SEQUENCE 383 AA; 41522 MW; 96BCD91F1B9AE7E CRC64;
 Query Match 39.4%; Score 43; DB 1; Length 383;
 Best Local Similarity 36.8%; Pred. No. 8.6;
 Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
 OY 1 TPDINPAMYXXRGIRPYGR 19
 DB 263 TPQRTAWTAIRHVDINO 281
 RESULT 8
 YAAJ_ECOLI STANDARD; PRT; 476 AA.
 ID YAAJ_ECOLI
 AC P30143;

DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL 51.7 KDA PROTEIN IN THRC-TALB INTERGENIC REGION (ORF8).
 GN YAAU.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_TaxId=562;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RC MEDLINE-92334977; PubMed-1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 0-2.4 min region."
 RT Nucleic Acids Res. 20:3305-3308(1992).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RC MEDLINE-97426617; PubMed-9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RT Science 277:1453-1474(1997).
 RL [2]
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE SODIUM:ALANINE SYMPORTER FAMILY
 (SAF). STRONG. TO H. INFLUENZAE H10183.
 CC (SAF).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: D10483; -; NOT_ANNOTATED_CDS.
 CC EMBL: AE000115; AAC73118.1; -;
 CC Ecogene: EGI1555; yaaU.
 DR InterPro: IPR001463; -;
 DR Pfam: PF01235; Na_Ala_symp; 1.
 DR PRINTS: PR00175; NAALASMPORF.
 DR PROSITE: PS00873; NA_ALANINE_SYMP; 1.
 KW Hypothetical protein; Transmembrane; Inner membrane; Transport;
 KM Symport.
 FT TRANSMEM 4 24 POTENTIAL.
 FT TRANSMEM 81 101 POTENTIAL.
 FT TRANSMEM 141 161 POTENTIAL.
 FT TRANSMEM 174 194 POTENTIAL.
 FT TRANSMEM 207 227 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 300 320 POTENTIAL.
 FT TRANSMEM 351 371 POTENTIAL.
 FT TRANSMEM 391 411 POTENTIAL.
 FT TRANSMEM 414 434 POTENTIAL.
 SO SEQUENCE 476 AA; 51662 MW; 2F6EB2E12E126E63 CRC64;

Query Match 39.4%; Score 43; DB 1; Length 476;
 Best Local Similarity 44.4%; Pred. No. 11;
 Matches 8; Conservative 2; Mismatches 2; Indels 6; Gaps 1;

OY 3 DIN-----PAWYXXRGI 14
 DB 120 DVNGGFRGGPAMTMARGL 137

RESULT 9

ARP_YEAST
 ID ARP_YEAST STANDARD; PRT; 719 AA.
 AC P32770; Q12228;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ARP PROTEIN.
 GN ARP OR ARP OR YDL167C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxId=4932;
 RN NCBI
 RP SEQUENCE FROM N.A.
 RC STRAIN-AH22;
 RC MEDLINE-93247548; PubMed-8483449;
 RA Wehner E.P., Rao E., Brendel M.;
 RT "Molecular structure and genetic regulation of SFA, a gene
 responsible for resistance to formaldehyde in Saccharomyces
 cerevisiae and characterization of its protein product."
 RT Mol. Gen. Genet. 237:351-358(1993).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RA Pohl T.M.;
 RT Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 RL [2]
 CC -1- SIMILARITY: CONTAINS 2 RANBP-TYPE ZINC FINGERS.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X68070; CAA48159.1; -;
 CC EMBL: Z67750; CAA81579.1; -;
 CC EMBL: Z74215; CAA98741.1; -;
 CC PIR: S31139; S31139.
 CC HSP: P04170; GRXN.
 CC SGD: S0002326; NRP1.
 CC InterPro: IPR000504; -;
 CC InterPro: IPR001876; -;
 CC Pfam: PF00076; rrm; 1.
 CC Pfam: PF00641; zf-RanBP; 2.
 CC PROSITE: PS00102; RRM; 1.
 CC PROSITE: PS00030; RRM_RNP_1; FALSE_NEG.
 KW Nuclear protein; zinc-finger; RNA-binding; Repeat.
 FT DOMAIN 226 322 RNA-BINDING (RRM).
 FT DOMAIN 490 564 ASN-RICH.
 FT CONFLICT 493 493 I -> N (IN REF. 1).
 SO SEQUENCE 719 AA; 79299 MW; ADA9BC09F582669 CRC64;

Query Match 39.4%; Score 43; DB 1; Length 719;
 Best Local Similarity 50.0%; Pred. No. 16;
 Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 7 AWYXXRGIRPVG 18
 DB 244 SWFTYGVVRPVG 255

RESULT 10
 ID Y762.METUA STANDARD; PRT; 342 AA.
 AC Q58172;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHETICAL PROTEIN M70762.

```

GN MJ0762.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus
NCBI_TaxID=2190;
RN
  11
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE-9633799; PubMed-8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
  jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0762 AND TO S.POMBE MALATE
  PERLEASE (MAE1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67521; AAB98753.1; -
DR TIGR: MJ0762; -
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 8 28
FT TRANSMEM 39 59
FT TRANSMEM 79 99
FT TRANSMEM 108 128
FT TRANSMEM 142 162
FT TRANSMEM 175 195
FT TRANSMEM 207 227
FT TRANSMEM 242 262
FT TRANSMEM 276 296
FT TRANSMEM 304 324
SQ SEQUENCE 342 AA; 39534 MW; 08FECE3E2C4955D8 CRC64;

Query Match 37.6%; Score 41; DB 1; Length 342;
Best Local Similarity 53.3%; Pred. No. 16;
Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 4 INPARYXXRGIRPVG 18
   :||| 11 11 11
DB 139 VNPQWY----IPVVG 149

RESULT 11
Y576_METJA STANDARD; PRT; 347 AA.
ID Y576_METJA
AC Q57996;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0576.
GN MJ0576.
OS Methanococcus jannaschii.
OC Archaea: Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;

```

```

RX MEDLINE-9633799; PubMed-8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Ullrich T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
  jannaschii."
RL Science 273:1058-1073(1996).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: STRONG, TO M.JANNASCHII MJ0762 AND TO S.POMBE MALATE
  PERLEASE (MAE1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U67506; AAB98567.1; -
DR TIGR: MJ0576; -
KW Hypothetical protein; Transmembrane; Transport.
FT TRANSMEM 15 35
FT TRANSMEM 46 66
FT TRANSMEM 84 104
FT TRANSMEM 111 131
FT TRANSMEM 149 169
FT TRANSMEM 182 202
FT TRANSMEM 214 234
FT TRANSMEM 249 269
FT TRANSMEM 283 303
FT TRANSMEM 312 332
SQ SEQUENCE 347 AA; 39556 MW; 632F671A1A1DE183 CRC64;

Query Match 37.6%; Score 41; DB 1; Length 347;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 2; Indels 4; Gaps 1;

QY 4 INPARYXXRGIRPVG 18
   :||| 11 11 11
DB 146 VNPQWY----IPVVG 156

RESULT 12
CYGR_ARBPV STANDARD; PRT; 986 AA.
ID CYGR_ARBPV
AC P11528;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RESACT RECEPTOR PRECURSOR (GUANYLATE CYCLASE) (EC 4.6.1.2).
OS Arabacia punctulata (Punctate sea urchin).
OC Echinoidae; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Echinodermata; Eleutherozoa; Echinozoa;
OX NCBI_TaxID=7641;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Testis;
RX MEDLINE-88318927; PubMed-2901039;
RX Singh S., Lowe D.G., Thorpe D.S., Rodriguez H., Kuang W.-J.,
RX Dangott L.J., Chinkers M., Goeddel D.V., Garbers D.L.;
RT "Membrane guanylate cyclase is a cell-surface receptor with homology
  to protein kinases."
RL Nature 334:708-712(1988).
CC -1- FUNCTION: IMPLICATED AS A CELL-SURFACE RECEPTOR ON SPERMATOZOEA
  FOR 'RESACT' A CHEMOTACTIC PEPTIDE, AND ON VARIOUS OTHER CELLS
  AS A RECEPTOR FOR ATRIAL NATRIURETIC PEPTIDE.
CC

```

```

CC -1- CATALYTIC ACTIVITY: GTP -> 3',5'-CYCLIC GMP + PYROPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- SIMILARITY: SOME SIMILARITY WITH CONSERVED REGION OF CATALYTIC
CC DOMAIN OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X12874; CAA31367.1; -.
CC PIR: S05480; OYRGA.
CC InterPro: IPR000719; -.
CC InterPro: IPR001828; -.
CC Pfam: PF01094; ANF_receptor; 1.
CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
CC Receptor: Transmembrane; Glycoprotein; Phosphorylation; Lyase;
CC GMP synthesis; Signal.
CC SIGNAL 1 21
CC CHAIN 1 22 986 RESACT RECEPTOR.
CC DOMAIN 22 507 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 508 528 POTENTIAL.
CC DOMAIN 529 986 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 568 836 PROTEIN KINASE LIKE.
CC CARBOHYD 185 185 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 410 410 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 986 AA; 111284 MW; B40238A74CCAF52 CRC64;

Query Match 37.6%; Score 41; DB 1; Length 986;
Best Local Similarity 62.5%; Pred. No. 46;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 PDINPAMY 9
Db 475 PDLNPMWH 482

RESULT 13
Y493.MYCTU STANDARD: PRT; 329 AA.
AC Q11158;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOHETICAL 35.4 KDA PROTEIN RV0493C.
GN RV0493C OR MTCY2069.19C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Mycobacterium.
OX NCBI_Taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekle A.F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
CC -1- SIMILARITY: SOME, TO M.LEPRAE B2168_F2_93.
CC -----

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z77162; CAB00954.1; -.
CC TubercuList; RV0493C; -.
CC Hypothetical protein.
CC SEQUENCE 329 AA; 35427 MW; BE4B524750277B56 CRC64;

Query Match 36.7%; Score 40; DB 1; Length 329;
Best Local Similarity 42.9%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 PDINPAMXXXXR 15
Db 90 PAAGPAMFDIAGVR 103

RESULT 14
CHOD_STRSQ STANDARD: PRT; 546 AA.
ID CHOD_STRSQ
AC P12676;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE CHOLESTEROL OXIDASE PRECURSOR (EC 1.1.3.6) (CHOD).
GN CHOA.
OS Streptomyces sp. (strain SA-COO).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_Taxid=1931;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=89123081; PubMed=2914858;
RA Ishizaki T., Hirayama N., Shinkawa H., Nimi O., Murooka Y.;
RT "Nucleotide sequence of the gene for cholesterol oxidase from a
RT Streptomyces sp."
RL J. Bacteriol. 171:596-601(1989).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS), AND MUTAGENESIS.
RA MEDLINE=99211873; PubMed=10194345;
RA Yue Q.K., Kass I.J., Sampson N.S., Vrielink A.;
RT "Crystal structure determination of cholesterol oxidase from
RT Streptomyces and structural characterization of key active site
RT mutants."
RL Biochemistry 38:4277-4286(1999).
CC -1- CATALYTIC ACTIVITY: CHOLESTEROL + O(2) -> CHOLEST-4-EN-3-ONE +
CC H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: CHOLESTEROL METABOLISM.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO THE GMC OXIDOREDUCTASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M31939; AAA26719.1; -.
CC PIR: A32260; A32260.
CC PDB: 1B4V; 06-JAN-99.
CC PDB: 1B8S; 09-FEB-99.
CC PDB: 1CBO; 10-MAR-99.
CC PDB: 1CC2; 11-MAR-99.

```

DR InterPro: IPR000172; -
 DR InterPro: IPR001167; -
 DR Pfam: PF01319; CHOD; 1.
 DR PROSITE: PS00623; GMC_OXRED_1; 1.
 DR PROSITE: PS00624; GMC_OXRED_2; FALSE_NEG.
 KW Oxidoreductase; Signal; Flavoprotein; FAD; 3D-structure.
 FT CHAIN 1 42
 FT NP_BIND 43 546 CHOLESTEROL OXIDASE.
 FT ACT_SITE 54 70 FAD (ADP PART) (POTENTIAL).
 FT ACT_SITE 398 398 PROTON ACCEPTOR.
 FT ACT_SITE 484 484
 FT MUTAGEN 398 398 E->Q: REDUCED ACTIVITY.
 FT MUTAGEN 484 484 H->N,Q: REDUCED ACTIVITY.
 SQ SEQUENCE 546 AA; 58993 MW; EF22A1PE5EA68D21 CRC64;

Query Match 36.7%; Score 40; DB 1; Length 546;
 Best Local Similarity 36.8%; Pred. No. 38;
 Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVGRF 20
 DB 98 PDKRSSMFKNRTREAPLGSE 116

RESULT 15
 POL_MLVK STANDARD; PRT; 581 AA.
 AC P31795;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE POLYPROTEIN [CONTAINS: PROTEASE (EC 3.4.23.-); REVERSE
 TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)] (FRAGMENT).
 GN POL.
 OS Radiation murine leukemia virus (strain Kaplan).
 OC Viruses; Retrovirdae; Retroviridae; Mammalian type C retroviruses.
 OX NCBI_TaxID=31689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92333703; PubMed=162969;
 RA Poliquin L., Bergeron D., Fortier J.L., Paquette Y., Bergeron R.,
 RA Rassart E.;
 RT "Determinants of thymotopism in Kaplan radiation leukemia virus and
 RT nucleotide sequence of its envelope region.";
 RL J. Virol. 66:5141-5146(1992).
 CC -1- PPM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-POL
 CC POLYPROTEIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE
 CC RETROPEPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M93052; AAA6525.1; -
 CC PIR: A42743; A42743.
 CC MEROPS: A02.008; -
 DR InterPro: IPR001037; -
 DR InterPro: IPR001584; -
 DR InterPro: IPR001969; -
 DR InterPro: IPR002156; -
 DR Pfam: PF00552; Integrase; 1.
 DR Pfam: PF00075; rnses; 1.
 DR Pfam: PF00665; rve; 1.
 DR PROSITE: PS00141; ASP_PROTEASE; PARTIAL.
 DR Hydrolase; Transferase; RNA-directed DNA polymerase;
 KW Aspartyl protease; Endonuclease; Polyprotein.

FT NON_TER 1 1
 SQ SEQUENCE 581 AA; 65157 MW; 8D7A38694C8E036E CRC64;

Query Match 36.7%; Score 40; DB 1; Length 581;
 Best Local Similarity 41.2%; Pred. No. 40;
 Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPINPAMYXXRGIRPV 17
 DB 556 TPPIRPSMRYQSRONPL 572

Search completed: April 17, 2001, 15:48:53
 JOD time: 541 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 17, 2001, 15:48:13 ; Search time 115.07 Seconds
(without alignments)
22.409 Million cell updates/sec

Title: US-09-446-543A-73
Perfect score: 109
Sequence: 1 TPDINPAMWXXRGIRPVGRFX 22

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP unclassified:*
13: SP vertebrate:*
14: SP virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	75.2	117	13 Q9W624	Q9W624 carassius a
2	49	45.0	790	10 Q9M371	Q9M371 arabidopsis
3	46	42.2	54	4 Q9UJF9	Q9UJF9 homo sapien
4	46	42.2	333	2 Q9PH76	Q9PH76 xyella fas
5	46	42.2	465	4 Q60687	Q60687 homo sapien
6	45	42.2	540	10 Q9L620	Q9L620 oryza sativ
7	45	41.3	767	5 Q20170	Q20170 caenorhabd
8	44	40.4	284	1 Q50128	Q50128 pyrococcus
9	44	40.4	419	4 Q9Y276	Q9Y276 homo sapien
10	43.5	39.9	1501	10 Q9SD86	Q9SD86 arabidopsis
11	43	39.4	232	2 Q9R221	Q9R221 deinococcus
12	43	39.4	309	5 Q17234	Q17234 caenorhabd
13	43	39.4	1296	5 Q22452	Q22452 caenorhabd
14	42.5	39.0	333	5 Q9RJ10	Q9RJ10 streptomyce
15	42.5	39.0	443	5 Q19879	Q19879 caenorhabd
16	42	38.5	293	2 Q53855	Q53855 spiroplasma
17	42	38.5	428	2 Q9PKF7	Q9PKF7 chlamydia m
18	41.5	38.1	345	2 Q9R8V2	Q9R8V2 bacillus ha
19	41	37.6	97	2 Q33440	Q33440 pseudomonas

20	41	37.6	128	2 P74747	P74747 synechocyst
21	41	37.6	284	1 Q9V0Q3	Q9V0Q3 pyrococcus
22	41	37.6	328	4 Q9WVR5	Q9WVR5 homo sapien
23	41	37.6	338	5 Q18729	Q18729 caenorhabd
24	41	37.6	343	2 Q9Z630	Q9Z630 streptomyce
25	41	37.6	414	2 Q33480	Q33480 proplionbac
26	41	37.6	547	2 Q55020	Q55020 streptomyce
27	41	37.6	2155	11 Q08523	Q08523 mus musculu
28	40.5	37.2	779	3 Q9P519	Q9P519 neurospora
29	40.5	37.2	1540	10 Q9SD90	Q9SD90 arabidopsis
30	40	36.7	76	5 Q9V8L0	Q9V8L0 drosophila
31	40	36.7	153	3 Q9T133	Q9T133 bacterioph
32	40	36.7	181	3 Q9USE8	Q9USE8 schizosacch
33	40	36.7	184	2 Q68838	Q68838 streptomyce
34	40	36.7	207	5 Q27455	Q27455 brachycentr
35	40	36.7	213	5 Q9VHT8	Q9VHT8 drosophila
36	40	36.7	244	10 Q9SD99	Q9SD99 arabidopsis
37	40	36.7	324	2 Q9Z554	Q9Z554 streptomyce
38	40	36.7	341	2 Q9X7Y1	Q9X7Y1 streptomyce
39	40	36.7	359	3 Q59676	Q59676 schizosacch
40	40	36.7	390	2 Q9PH18	Q9PH18 xyella fas
41	40	36.7	402	2 Q9KIW3	Q9KIW3 chiodaciliu
42	40	36.7	430	1 Q27142	Q27142 mechanobact
43	40	36.7	462	10 Q39249	Q39249 arabidopsis
44	40	36.7	466	2 Q54042	Q54042 pseudomonas
45	40	36.7	562	5 Q9VYN4	Q9VYN4 drosophila

ALIGNMENTS

RESULT 1
Q9W624 PRELIMINARY: PRT; 117 AA.
ID Q9W624;
AC Q9W624;
DR 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE C-RF AMIDE PRECURSOR.
OS Carassius auratus (Goldfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprininae; Carassius.
OX NCBI_TaxID=7957;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RT "Carassius Rfamade (C-RF amide).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB020024; BAA76662.1; --
SQ SEQUENCE 117 AA; 12879 MW; DSDC4CB22038C2B0 CRC64;

Query Match 75.2%; Score 82; DB 13; Length 117;
Best Local Similarity 60.0%; Pred. No. 3.3e-06;
Matches 12; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Qy 1 TPDINPAMWXXRGIRPVGRFX 20
Db 56 SPEIDPFVWVGVRPIGRFX 75
RESULT 2
Q9M371 PRELIMINARY: PRT; 790 AA.
ID Q9M371;
AC Q9M371;
DR 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL 87.4 KDA PROTEIN.
GN F15G16.60.
OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN
 RP SEQUENCE FROM N.A.
 RA De Haan M., Meares A.C., Grivell L.A., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quelier F., Salanoubat M.,
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN
 [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL132959; CAB71097.1; -
 KW Hypothetical protein.
 SO SEQUENCE 790 AA; 87375 MW; B222724B75690F30 CRC64;

Query Match 45.0%; Score 49; DB 10; Length 790;
 Best Local Similarity 47.4%; Pred. No. 7.4;
 Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGIRPVGRF 20
 DB 366 PPHNPRTYSGRGLQPHGRW 384

RESULT 3
 OY09UF9 PRELIMINARY; PRT; 54 AA.
 AC O9UF9;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE DJ479J7.3 (SUSHI-REPEAT PROTEIN (SRPUL)) (FRAGMENT).
 GN DJ479J7.3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lawlor S.;
 RL Submitted (SEP-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL035608; CAB55682.1; -
 FT NON_TER 54
 SO SEQUENCE 54 AA; 6110 MW; E2F3C39F7B961A9F CRC64;

Query Match 42.2%; Score 46; DB 4; Length 54;
 Best Local Similarity 43.8%; Pred. No. 1.3;
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

OY 1 TPDPINPAMYXXRGIRP 16
 DB 18 TPAYVPTPTWYAGSGYYP 33

RESULT 4
 OY09PH76 PRELIMINARY; PRT; 333 AA.
 AC O9PH76;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE HYDROXYBENZONATE OCTAPRENYLTRANSFERASE.
 GN XF0068.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin L., Bove J.M., Brites M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincini A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frome M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurame E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E.Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A.Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Zeldanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-157(2000).
 DR EMBL; AE003860; AAF82881.1; -
 DR INTERPRO; IPR000537; -
 DR PFM; PF01040; UBLA: 1.
 DR PROSITE; PS00943; UBLA: UNKNOWN_1.
 SO SEQUENCE 333 AA; 37931 MW; ECF3F4716C962B95 CRC64;

Query Match 42.2%; Score 46; DB 2; Length 333;
 Best Local Similarity 53.3%; Pred. No. 9.2;
 Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 4 INPAMYXXRGIRPVG 18
 DB 54 LDPYKRLARGDRPVG 68

RESULT 5
 OY060687 PRELIMINARY; PRT; 465 AA.
 AC O60687;
 DT 01-AUG-1998 (TREMBlrel. 07, Created)
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
 DE 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE SUSHI-REPEAT PROTEIN.
 GN SRPUL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kurosawa H., Inukai T., Inaba T., Gol K., Chang K.-S., Shinjo T.,
 RA Rakestraw K.M., Naeye C.W., Look T.A.,
 RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF060567; AAC15765.1; -
 DR INTERPRO; IPR000436; -
 DR INTERPRO; IPR001128; -
 DR PFM; PF00084; sush1: 3.
 DR PROSITE; PS00086; CYTOCHROME_P450; UNKNOWN_1.
 SO SEQUENCE 465 AA; 52971 MW; 4D752B187FF3EFB8 CRC64;

Query Match 42.2%; Score 46; DB 4; Length 465;

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=99097350; PubMed=9878253;
 RA Petruzzella V., Tiranti V., Fernandez P., Ianna P., Carozzo R.,
 Zeviani M.;
 RT "Identification and characterization of human cDNAs specific to BCS1,
 PPT12, SCOT, COX15, and COX11, five genes involved in the formation
 and function of the mitochondrial respiratory chain.";
 RL Genomics 54:494-504(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=96207227; PubMed=8619474;
 RA Andersson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;
 RT "A 'double adaptor' method for improved shotgun library
 construction.";
 RL Anal. Biochem. 236:107-113(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RX MEDLINE=97264341; PubMed=9110174;
 RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
 Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
 RT "Large-scale concatenation cDNA sequencing.";
 RL Genome Res. 7:353-358(1997).
 DR EMBL: AF026849; AAD08638.1; -;
 DR EMBL: AF038195; AAB97365.1; -;
 DR INTERPRO: IPR001939; -;
 DR PFM: PF00004; AAA; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 419 AA; 47534 MW; 7F0F98BA62F2CB8 CRC64;

Query Match 40.4%; Score 44; DB 4; Length 419;
 Best Local Similarity 70.0%; Pred. No. 25;
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 5 NPAMYXXRGI 14
 DB 211 NPKMYTDRCI 220

RESULT 10
 O9SD86 PRELIMINARY; PRT; 1501 AA.
 AC O9SD86;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 GN HYPOTHETICAL 164.4 KDA PROTEIN.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Van Der Schueren J., Chung Y.J., Voet M., Robben J.,
 Volckaert G., Bancroft I., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL133421; CAB62610.1; -;
 KW Hypothetical protein.
 SO SEQUENCE 1501 AA; 164360 MW; EAD2AE0C9E8E5D2 CRC64;

Query Match 39.9%; Score 43.5; DB 10; Length 1501;

Best Local Similarity 34.5%; Pred. No. 1.2e+02;
 Matches 10; Conservative 5; Mismatches 5; Indels 9; Gaps 2;
 OY 1 TPDINPAMYXXRGI-----IRPV---GRF 20
 DB 1184 SPQMAPSWISQYGFKNGLVQVNDIGRF 1212

RESULT 11
 O9R221 PRELIMINARY; PRT; 232 AA.
 AC O9R221;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 GN HYPOTHETICAL 25.0 KDA PROTEIN.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Mofatt K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fetschman R.D.,
 Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 Fraser C.M.;
 RT "Genome Sequence of the Radioreistant Bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL: AE001862; AAF12317.1; -;
 DR TIGR: DRA0132; -;
 KW Hypothetical protein.
 SQ SEQUENCE 232 AA; 24979 MW; A044FA2F38435DA7 CRC64;

Query Match 39.4%; Score 43; DB 2; Length 232;
 Best Local Similarity 46.7%; Pred. No. 20;
 Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

OY 2 PDINPAMYXXRGI 16
 DB 130 PDHRAAMHLRGVLP 144

RESULT 12
 O17234 PRELIMINARY; PRT; 309 AA.
 AC O17234;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 GN K10F12.4 PROTEIN.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnson L.,
 Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,

RA Smaidon N., Smith A., Sonnhamer E., Staden R., Sulston J.,
 RA Thierly-Meg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Woldmann P., Beck C.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF025462; AAB71002.1; -;
 DR INTERPRO: IPR000521; -;
 DR PFAM: PF00043; GST; 1;
 SQ SEQUENCE 309 AA; 35021 MW; FA2091931D8AD2DC CRC64;

Query Match 39.4%; Score 43; DB 5; Length 309;
 Best Local Similarity 44.4%; Pred. No. 27;
 Matches 8; Conservative 3; Mismatches 5; Indels 2; Gaps 1;
 Oy 2 PDINPAMYXXRGIRPVGR 19
 ||:||||:||||
 Db 131 PDSPNMLPKS--PIGR 146

RESULT 13

022452 PRELIMINARY; PRT; 1296 AA.

AC 022452;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
 DE SIMILAR TO AGRIN AND FOLLISTATIN.
 GN T13C2.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitidae;
 OC Rhabdilitidae; Peleoderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Anscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Keshava J., Kirsten J., Laister N., Latreille P.,
 RA Lighthouse J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Showkneen R.,
 RA Smaidon N., Smith A., Sonnhamer E., Staden R., Sulston J.,
 RA Thierly-Meg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP Du Z.;
 RA Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U40030; AAB1133.1; -;
 DR HSP: P37109; 1PCE.
 DR INTERPRO: IPR000561; -;
 DR INTERPRO: IPR001239; -;
 DR INTERPRO: IPR001791; -;

DR INTERPRO: IPR002049; -;
 DR INTERPRO: IPR002350; -;
 DR PFAM: PF00050; Kazal; 9.
 DR PFAM: PF00053; laminin_EGF; 2.
 DR PFAM: PF00054; laminin_G; 1.
 DR PRINTS: PR00290; KAZALINBTR.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
 SQ SEQUENCE 1296 AA; 145178 MW; 05094BC185836990 CRC64;

Query Match 39.4%; Score 43; DB 5; Length 1296;
 Best Local Similarity 45.0%; Pred. No. 1.2e+02;
 Matches 9; Conservative 2; Mismatches 7; Indels 2; Gaps 1;
 Oy 1 TPDINPAMYXXRGIRPVGR 20
 |||||:|:|:
 Db 975 TPDINSDWYFSR--KDINRF 992

RESULT 14

09RJ10 PRELIMINARY; PRT; 333 AA.

AC 09RJ10;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE HYPOTHETICAL 36.3 KDA PROTEIN.
 GN SCF73.06C.
 OS Streptomyces coelicolor.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Seeger K.J., Harris D.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Parkhill J., Barrell B.G., Rajandream M.A.;
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2);
 RA Redenbach M., Kleser H.M., Denapate D., Elchner A., Cullum J.,
 RA Kinsash H., Hopwood D.A.;
 RT "A set of ordered cosmid and a detailed genetic and physical map for
 the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
 RL Mol. Microbiol. 21:77-96(1996).
 DR EMBL: AL121746; CAB57411.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 333 AA; 36312 MW; 884FF2861837FD76 CRC64;

Query Match 39.0%; Score 42.5; DB 2; Length 333;
 Best Local Similarity 32.3%; Pred. No. 35;
 Matches 10; Conservative 2; Mismatches 8; Indels 11; Gaps 1;
 Oy 1 TPDINPAM-----YXXRGIRPVGR 20
 |||||:|:|:
 Db 19 TPWEPPAWTEAIGWTEARLAHGLRPTGRW 49

RESULT 15

019879 PRELIMINARY; PRT; 443 AA.

AC 019879;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

```

DE F28D1.8 PROTEIN.
GN F28D1.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea.
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Baynes C.;
RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-94150718; PubMed-7906398;
RA Wilson R., Alnscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaiden N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P., III of C.
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL; 270684; CAA94603.1;
SQ SEQUENCE 443 AA; 48865 MW; D9137BACCE575E CRC64;

Query Match 39.0%; Score 42.5; DB 5; Length 443;
Best Local Similarity 37.5%; Pred. No. 47;
Matches 9; Conservative 3; Mismatches 3; Indels 9; Gaps 1;

OY 4 INPWWYXXR-----GIRPVG 18
DB 53 INPSWEDWRSVSIQGHGIMPIG 76

```

Search completed: April 17, 2001, 15:48:13
Job time: 566 sec